



Cite this article: Bauer R, Kaiser M. 2017
Nonlinear growth: an origin of hub
organization in complex networks. *R. Soc.
open sci.* **4**: 160691.
<http://dx.doi.org/10.1098/rsos.160691>

Received: 12 September 2016

Accepted: 21 February 2017

Subject Category:

Biology (whole organism)

Subject Areas:

biocomplexity/systems
biology/computational biology

Keywords:

complex networks, hubs, neural development,
physical space, rich-club organization,
nonlinear growth

Author for correspondence:

Marcus Kaiser
e-mail: m.kaiser@newcastle.ac.uk

Electronic supplementary material is available
online at [https://dx.doi.org/10.6084/m9.
figshare.c.3711967](https://dx.doi.org/10.6084/m9.figshare.c.3711967).

Nonlinear growth: an origin of hub organization in complex networks

Roman Bauer^{1,2} and Marcus Kaiser^{1,2}

¹Institute of Neuroscience, Newcastle University, Newcastle upon Tyne NE2 4HH, UK

²Interdisciplinary Computing and Complex BioSystems Research Group (ICOS), School
of Computing Science, Newcastle University, Newcastle upon Tyne NE1 7RU, UK

RB, 0000-0002-7268-9359

Many real-world networks contain highly connected nodes called hubs. Hubs are often crucial for network function and spreading dynamics. However, classical models of how hubs originate during network development unrealistically assume that new nodes attain information about the connectivity (for example the degree) of existing nodes. Here, we introduce hub formation through nonlinear growth where the number of nodes generated at each stage increases over time and new nodes form connections independent of target node features. Our model reproduces variation in number of connections, hub occurrence time, and rich-club organization of networks ranging from protein–protein, neuronal and fibre tract brain networks to airline networks. Moreover, nonlinear growth gives a more generic representation of these networks compared with previous preferential attachment or duplication–divergence models. Overall, hub creation through nonlinear network expansion can serve as a benchmark model for studying the development of many real-world networks.

1. Introduction

1.1. Background

One of the most commonly used network growth models that generate hubs [1,2], which often are essential for network function [3,4] and spreading dynamics [5], is the preferential attachment model ('rich gets richer') [6]. However, it assumes that connections of a new node form depending on how well connected a potential target node is compared to all other existing nodes. The knowledge of such global information is often difficult to justify in a biological context; for example, axonal growth cones in neuronal networks can only sense their local environment. It is unclear whether and how the preferential attachment model can be applied to such network growth dynamics, as it often cannot provide a mechanistic explanation. The duplication–divergence model [7], which is inspired by the evolution of protein–protein

interaction (PPI) networks, on the other hand, needs a duplication of previously existing connection patterns that is unlikely for many non-biological networks (e.g. flight connections between airports). Moreover, the divergence step entails a non-trivial coordination between duplicated nodes (i.e. a new connection is pruned with a pre-specified probability, but only if its counterpart is not). Such communication is difficult to justify in many biological networks, where the information exchange between maturing connections is limited.

1.2. The principle of nonlinear growth

We propose a novel model that is based on exponential network growth. It assumes that the number of new nodes increases nonlinearly with time, which is a general principle of growth and development: systems ranging from the worldwide web to neural networks in the brain experience episodes of exponential growth where the number of new nodes at one time will be larger than the number of new nodes at a previous time. As an example for growth in biological systems, the number of cells per unit volume (c) in the growth phase of bacterial cultures can be described by an exponential function: $c(n) = c_1 \cdot 2^n$, where c_1 is the initial unit volume of the culture, and n is the number of divisions a cell has undergone [8]. For brain evolution, it was proposed that new neural structures form by separation of already existing areas [9], with the number of brain areas then increasing exponentially. The early growth of the Internet [10] would be an example for non-biological systems.

2. Material and methods

2.1. Data analysis

Analysis and implementation of growth models were conducted using MATLAB R2014b (Mathworks Inc.). Visualization of the hub occurrences and node degrees across development (figure 1; electronic supplementary material, figure S4) was done by binning the maturation times, i.e. the times of nodes formation during network development. For the *Caenorhabditis elegans* and AIR datasets, three maturation bins (columns) were computed such that the number of nodes in each bin were approximately the same. For the PPI and macaque datasets, three age categories were assigned, and so the number of maturation bins is matched accordingly.

The results shown in figure 2 and the electronic supplementary material, figure S2A are computed based on 200 networks for each of the parameter sets. Random, regular networks of equal size were created as follows: first, the mean edge density of these nonlinearly grown networks was computed. Then, the corresponding probability of edge formation in random, regular network growth was computed (this regular growth corresponds to the N_L model with parameter $d=1$). Using this probability, 200 regular, linearly grown networks were generated for each parameter set. Finally, the corresponding coefficient of variation (CV) and hub-rich-club coefficient (HRCC) values were computed, which served as the comparative values for the shown relative deviations. Owing to high demands in computer resources, the results shown in the electronic supplementary material, figure S2B are based on 20 networks.

For the assessment of the model performances, 1000 network samples were generated for each analysed network measure (CV value, HRCC measure and the maturational pattern of hub occurrence). Network samples where the number of edges deviated by more than 10% from the target number were replaced by newly generated samples matching this criterion. For the CV and HRCC values, a model was classified as explanatory if a statistical test is passed for the experimental value to stem from the model-generated distribution, using the z-statistic with a significance of 1%, or if the value derived from the respective dataset was within the model-generated interval. For the maturational trajectories, the same approach was taken for the percentage of hubs in each maturation bin. A model was classified as explanatory if at least one of the model-generated numbers of hubs matched the observed value, or if the model-generated distribution passed the z-test for the observed value.

Rich-club coefficients and complementary network measures (clustering coefficient C , modularity index Q and characteristic path length L) were computed using the MATLAB scripts obtained from the Brain Connectivity Toolbox [11] (<https://sites.google.com/site/bctnet/measures/list>). Degree preserving reference networks were computed using the *randmio_und.m* function (also from the Brain Connectivity Toolbox).

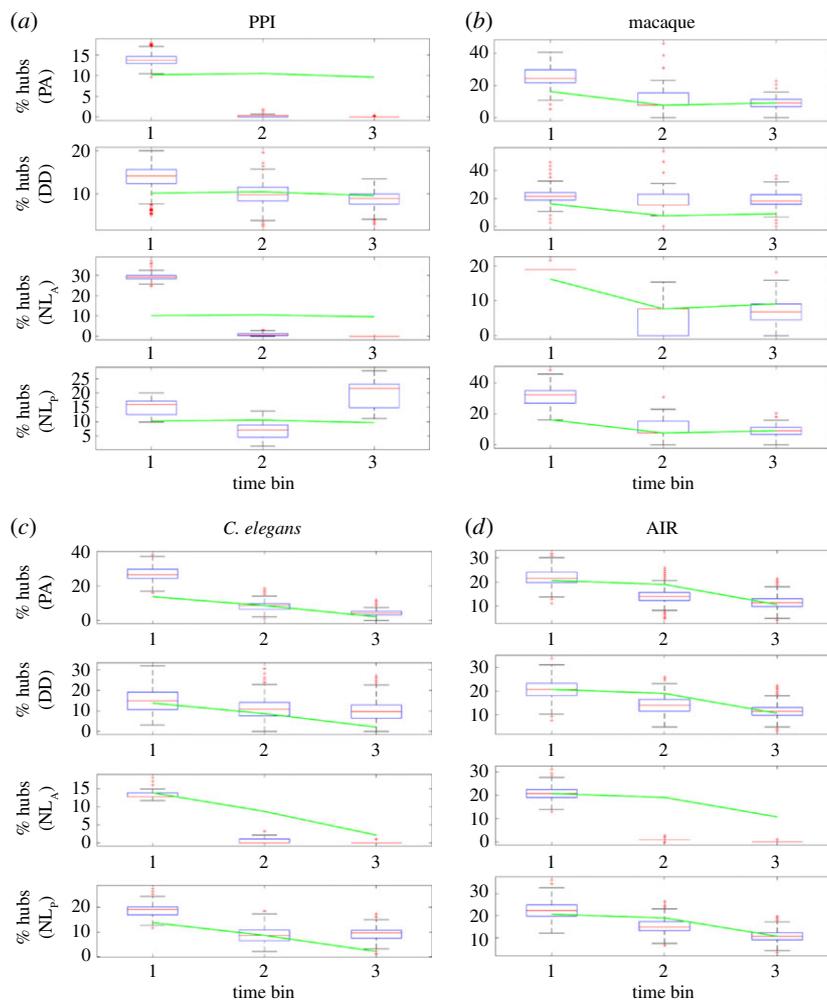


Figure 1. Trajectories of hub occurrences for the PPI (a), macaque (b), *C. elegans* (c) and AIR (d) datasets. Green lines indicate the experimental data, while the box-plots show the results of the (optimized) model-generated samples. The overall performance of the models is summarized in table 1. The green lines are identical for the same datasets, but are displayed within different axes.

2.2. Parameter optimization

We conducted parameter optimization using simulated annealing, such that the model-generated networks exhibit measures as close as possible to the empirical values. The MATLAB function `simulannealbnd.m` was used for conducting simulated annealing. The cost function was defined as: $A \cdot r_1 + r_2$. r_1 and r_2 are the relative errors of the number of edges and the CV or HRCC values, respectively. A is a penalty factor, and was manually adjusted in case the generated mean number of edges deviated largely from the empirical value (electronic supplementary material, figure S3). For the assessment of the models' performance, model-generated networks where the number of edges was beyond 10% of the dataset's edges were discarded. The resulting distribution of the number of edges had to pass a test using the z-statistic on the 1% significance level. In total, 1000 networks were generated for the visualizations of the CV values, the HRCC measures and the percentage of hub nodes.

Simulated annealing was also used for optimizing model parameters to yield the percentages of hubs in the three maturation bins. This was done under the null hypothesis that the observed percentage of hubs in a particular maturation bin stems from the model-generated distribution. The probability for observing the measured percentage under this null-hypothesis, as obtained from a z-test, was maximized in order to obtain the model parameters.

As the PPI and AIR datasets are bidirectional, we implemented that the formation of a connection in one direction also entails the connection into the other. Therefore, the target number of edges used in the optimization process was half the actual edges. The model-generated connectivity matrices were added to their transposed form, such that the resulting matrix was symmetrical and ideally

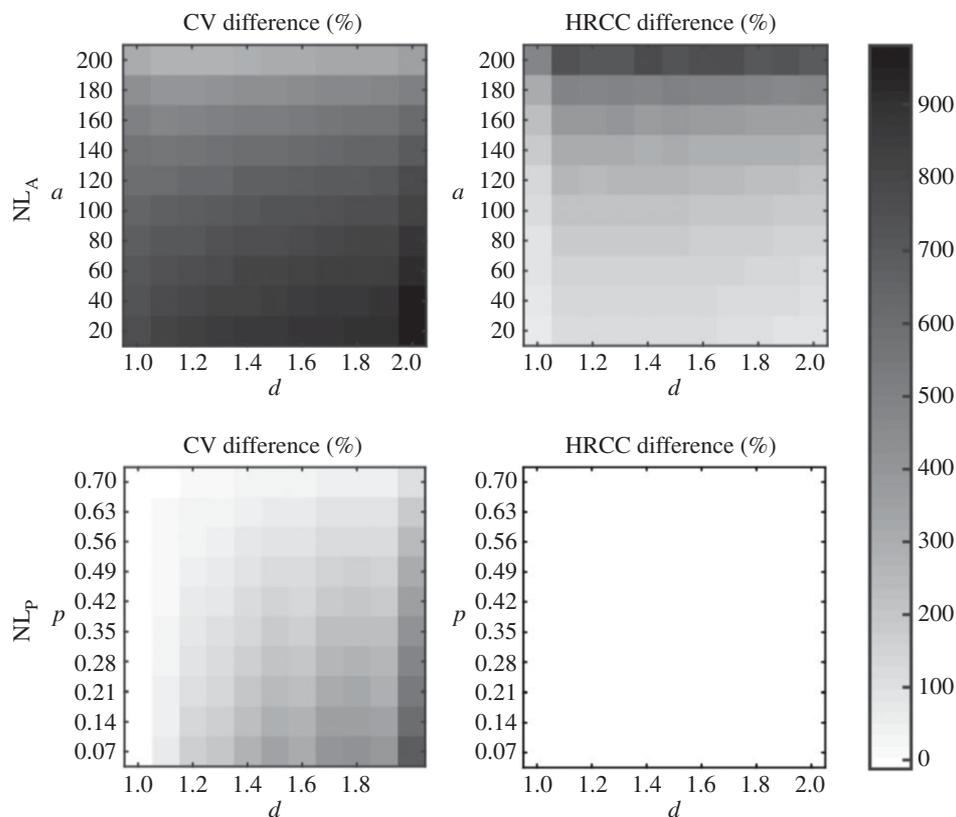


Figure 2. Influence of model parameters on degree variation (CV) and mean rich-club coefficient between hubs (HRCC) for the NL_A and the NL_P models, across different model parameters. The greyscale heatmap displays relative differences from CV and HRCC values of regular, random networks with the same number of nodes and edges. Darker shadings indicates higher CV and HRCC values, respectively, than expected from the control networks. Overall, the figure demonstrates that nonlinear growth can produce hub-related complex network properties. In contrast with the NL_A model, the NL_P model does not generate networks where hubs preferentially connect with each other (for all parameter constellations shown, a Wilcoxon rank sum test at the 5% significance level failed to reject the null hypothesis that the deviations have zero median). This is because the early-born nodes are equally as likely to connect to other early-born nodes as to later-born nodes, because the connection probability is fixed throughout network growth. Analogous results for the PA and DD models and larger networks are shown in the electronic supplementary material, figure S2.

matched the actual number of edges of the two datasets. For the assessments with regard to the macaque and *C. elegans* datasets, the model-generated networks were generated by probabilistically adding bidirectional connections, in order to match the proportion of unidirectional versus bidirectional connections derived from the datasets.

3. Results

3.1. Datasets

In order to compare the explanatory power of the nonlinear growth model with other models of network growth, we analysed four datasets including longitudinal information (figure 3*a*). The collected datasets of evolving networks include (i) the protein–protein interaction network in yeast [12], (ii) axonal connections between brain regions in the rhesus monkey (macaque) [13], (iii) the network between neurons in the nematode *C. elegans* [2,14] and (iv) the network of flights between international airports worldwide [15]. The maturation time of a node (i.e. the time when the node is added to the growing network) is defined as (i) the time during evolution when a protein first occurs, (ii) the maturation time of a brain region during development, (iii) the birth time of a neuron and (iv) the year when an airport was established, respectively (cf. table 2). The datasets (ii) and (iv) were collected by us and are therefore novel. All the used datasets are described in more detail in the electronic supplementary material.

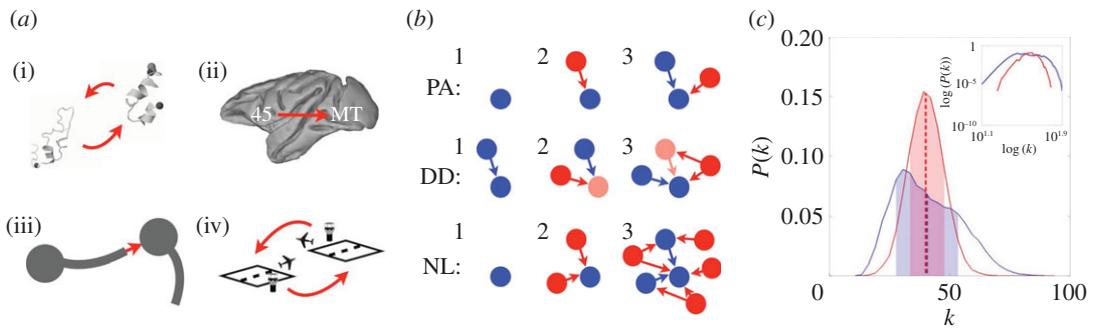


Figure 3. Datasets and models. (a) Types of connections in real-world networks: (i) interactions between proteins, (ii) connections between brain regions in the macaque, (iii) connections using chemical synapses, in *C. elegans*, and (iv) flight connections between airports. Note that connections for (ii) and (iii) may be unidirectional. (b) Growth models leading to highly connected nodes. PA, preferential attachment [6]: new nodes (red) preferentially connect to nodes with higher degrees. DD, duplication–divergence model: at each step, a random node (light red) is duplicated (red) together with its links. NL, nonlinear growth: the number of new nodes that are added at each step increases nonlinearly over time. New nodes project to already present nodes (blue) establishing on average a connections (NL_A) or link to each existing node with a probability p (NL_P). (c) Exemplary distributions for node degree k (inset: log–log plot) for networks from linear (red) and nonlinear growth (blue). Linear growth corresponds to the scenario where the network size increases linearly, i.e. only one node is added at each time step. Shaded areas show the standard deviation around the mean degree (dashed line). Nonlinear growth yields a wider distribution with more hubs.

3.2. The nonlinear growth model

The nonlinear growth model (NL) assumes that the network size increases nonlinearly/exponentially with time. At each developmental stage t , the network expands by d^t nodes (d is a parameter of the model) until a given network size is reached (figure 3b). The (rounded) number of newly formed nodes project to the already present nodes. Importantly, in our model the formation of connections does not rely on any properties of the nodes. Hence, a newly developed node does not need to check whether its target node fulfils specific conditions, but forms connections dependent upon a model parameter that is initially specified. NL substantially differs from the well-known Erdős–Rényi model (ER) [16], where the number of connections increases linearly, but the nodes are initially present already. As demonstrated in [17], linear network growth (i.e. addition of one node at a time), in addition to ER-like formation of connections, can yield complex network features such as small-worldness and scale-free degree distribution. Hence, the addition of new nodes is a crucial factor in network growth models. In this study, we show that also the dynamical nature of node addition strongly shapes the resulting network structures.

We distinguish two different scenarios of nonlinear growth. In the first case (NL_A), the number of target nodes that a new node projects to is computed as follows: a random value is drawn from a normal distribution with mean a and standard deviation $a/2$, where a indicates a model parameter. This value is rounded and limited by the cut-off points 0 and the current number of nodes in the network. Hence, the number of projections that a new node makes stays approximately the same throughout development. In a second scenario (NL_P), new nodes connect to any existing node with probability p , so that the number of a new node's outgoing projections is proportional to the number of existing nodes. In this scenario, the number of outgoing connections per node will increase during development, which is in contrast with the NL_A scenario. A simplified, algorithmic description of nonlinear growth and its two versions is shown in algorithm 1. As the NL model does not require any knowledge of a global variable, both scenarios are based solely on local information.

3.3. Alternative models

We have compared the NL model with the extended preferential attachment (PA) and the duplication–divergence (DD) models (figure 3b). For the classical preferential attachment model [6], at each time step a single node is added to the network and forms connections with existing nodes following a preference for highly connected nodes (i.e. hubs). For simplicity, our PA growth model starts from a single node (we confirmed that this initial condition yields scale-free degree distributions). The first parameter p_1 of the PA model determines to how many nodes the newly added node connects on average. A second

Algorithm 1

```

1: NonlinearGrowth( $N_{\text{Max}}, d, a$  or  $p$ )
2:
3:
4:  $t = 0$ 
5:  $n = 1$ 
6: while ( $n < N_{\text{Max}}$ ) do
7:    $t = t + 1$ 
8:    $n_{\text{New}} = \text{round}(d^t)$ 
9:   if  $n_{\text{New}} > N_{\text{Max}} - n$  then
10:     $n_{\text{New}} = N_{\text{Max}} - n$ 
11:   for  $i = 1 : n_{\text{New}}$  do
12:     NLA scenario:
13:     connect node  $i$  to  $\tilde{a}$  randomly selected, earlier born
14:     nodes (with  $\tilde{a} \sim \text{round}(\mathcal{N}(a, \frac{a^2}{4}))$  and  $\tilde{a} \geq 0$ )
15:     NLP scenario:
16:     connect node  $i$  to any earlier born node with probability  $p$ 
17:    $n = n + n_{\text{New}}$ 
18: return connectivity matrix

```

parameter p_2 of this extended version specifies a baseline probability to connect to any of the nodes in the network, independent of their degree. The probability that a newly added node connects to node i of the network is then computed as $p_i = p_1(k_i/\sum_j k_j) + p_2$. The PA model can be understood as an extension of a linear network growth model, where newly added nodes have knowledge about target nodes.

The DD model is implemented as described in [7]. Each time step consists of a duplication and a divergence step. In the duplication step, a random node is duplicated, i.e. it projects to the same connection partners as the original node. The connection between the two duplicated nodes prevails with probability p . In the subsequent divergence step, one of the two connections of the partners with the duplicated nodes is pruned with probability q . Maximally one of the two connections can be pruned, and so this model is based upon non-trivial coordination between the two connections. To allow for a fair comparison, both the alternative and the nonlinear growth models are based on two parameters.

3.4. Analysis of model-generated networks

We have assessed the networks' range of degrees through the coefficient of degree variation (CV), which is the standard deviation divided by the mean. For a linearly growing network with random formation of connections, all degrees are relatively close to the average degree, while for networks that contain highly connected nodes, the standard deviation of the node degree is larger due to such outlier nodes (figure 3c; electronic supplementary material, figure S1). We find that CV value distributions of model-generated networks are sufficiently narrow to allow to differentiate the models from each other, and so are well suited for a comparison of the different models. Moreover, because this study focuses on the development of hub-related network organization, we did not investigate other metrics of the degree distribution that are not directly related to hubs.

Complementary to the degree distribution, we also investigated the rich-club organization, which captures the tendency for hubs to interconnect with one another, rather than with low-degree nodes [18]. In many real-world networks, such as the neuronal connectivity of *C. elegans* [19] and the inter-areal projections in cat [20], macaque [21] or human [22], hub nodes connect with each other more strongly than with nodes of lower degree. Conversely, it has been shown that protein interaction networks do not contain rich clubs [18]. To quantify this tendency for connectivity among hubs, we used the rich-club coefficient [18] defined as $\phi(k) = 2E_{>k}/N_{>k}(N_{>k} - 1)$, where $E_{>k}$ is the number of edges among the $N_{>k}$ nodes that have a higher degree than k . Essentially, this coefficient measures the ratio of existing connections to the potential number of connections among a given set of nodes. Hence, the rich-club coefficient is a vector of values (different coefficients for different values of k). In order to simplify the comparison between different complex networks, we summarize the rich-club organization as the mean of the rich-club coefficients for all degrees k that qualify a node as being a hub. We define this so-called HRCC as: $\text{HRCC} = \sum_{k \in \Gamma} \phi(k)/|\Gamma|$, where Γ is the set of the degrees of all hub nodes in a given network

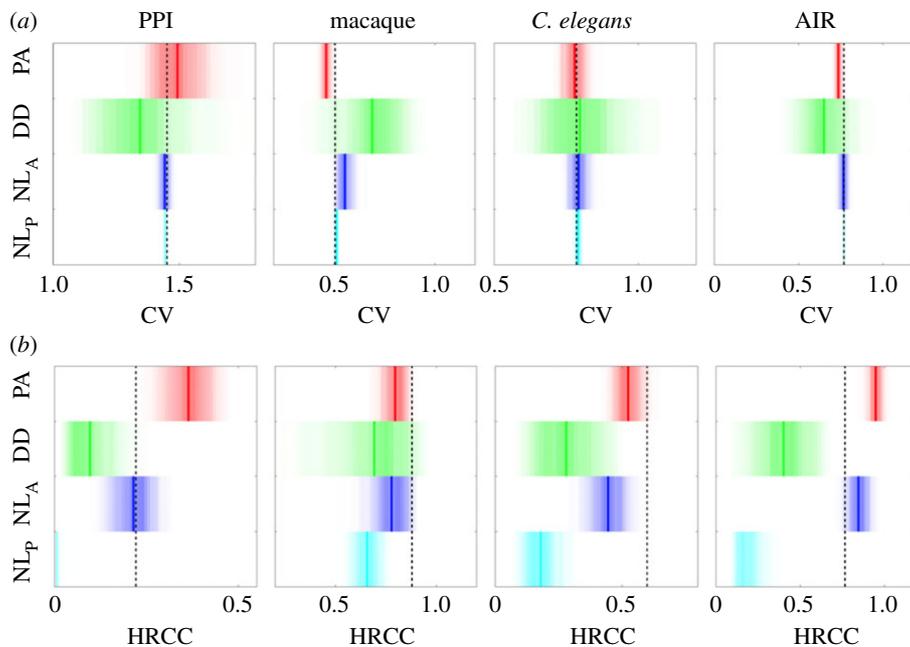


Figure 4. Generation of network properties. Degree variation (CV, *a*) and mean rich-club coefficient between hubs (HRCC, *b*) of the models in relation to values of real-world networks (black dashed lines). Red, green, blue and cyan shaded regions indicate the distribution of values that PA, DD, NL_A and NL_P could yield, respectively. Thick coloured lines indicate the mean values of the distributions. The respective, model-generated connection numbers are displayed in the electronic supplementary material, figure S3.

and $|I|$ is the number of elements in this set. We classify a node as a hub if the number of its connections k is at least one standard deviation above the average number of connections of a node ($k \geq \mu + \sigma$). Notably, rich-club organization is different from the assortativity property [18], and is a particularly well-esteemed measure in many biological networks.

Figure 2 and the electronic supplementary material, figure S2 visualize that nonlinear growth can indeed produce networks with CV and HRCC values that are much larger than those of regular, random networks across different parameter values. As expected, these hub-related network measures are very dependent on the two model parameters: the nonlinear growth exponent d strongly affects CV values, by increasing the spread of the degree distribution. The NL_P model yields, due to equal probability of connections among all nodes, no rich-club organization. However, for $d > 1$ it produces hubs across a wide spectrum of parameters. For the NL_A model, the CV values and the rich-club organization express themselves much beyond what is expected based on comparable, regular networks (note also the sudden increase of HRCC values for $d > 1$). Detailed and asymptotic behaviour of the network measures for the NL_P model are consistent across different network sizes (electronic supplementary material, figure S2). In addition to the CV and HRCC measures, the hub nodes were additionally analysed for their developmental time, and the models were assessed for their explanatory power with regard to this temporal aspect (figure 1).

We find that the analysed models reproduce many properties of the collected networks; preferential attachment matches some of the network measures (figure 4) but, for example, fails to generate the same number of connections when producing the CV value of the airline network (electronic supplementary material, figure S3). DD and NL models match both CV and number of connections. While the NL_P can account for all the CV values and the maturational trajectories, it cannot yield the HRCC values. Hub nodes are predominantly established early on [2,23] and show a distinct pattern of creation over time (electronic supplementary material, figure S4). Other relevant developmental properties are shown in the electronic supplementary material, figure S5.

Our results demonstrate that only the NL_A model can account for all the datasets' CV and HRCC values. It provides an intuitive explanation for the origins of rich-club connectivity; during early network development, there are not many nodes to project to, therefore most early born nodes project to each other. As early developing nodes are likely to become highly connected, hubs are predominantly connected among each other. Note, however, that nonlinear growth is also crucial: even a modified PA model, where the number of outgoing connections is also limited, cannot reproduce CV and HRCC

Table 1. Comparison of model performances. Agreement of preferential attachment (PA), duplication–divergence (DD) and nonlinear growth models (NL_A or NL_P) with features of real-world networks. Symbols denote whether hub occurrence time (\square), CV ($()$), HRCC ($-$) or all of the previous three (\boxplus) matched real-world networks.

	PA	DD	NL_A	NL_P
protein–protein interactions		\boxplus	+	\square
macaque cortical network	\boxplus	\boxplus	+	\square
<i>C. elegans</i> neuronal network	+	\square	+	\square
airport flight connections	\square	\boxplus	+	\square

Table 2. Properties of the collected datasets. The number of nodes, number of edges, the average degree μ , the maximum degree, the coefficient of variation (CV) and the hub-rich-club coefficient (HRCC) are listed.

	no. nodes	no. edges	μ	max(k)	CV	HRCC
protein–protein interactions	1857	9792	10.5	168	1.45	0.22
macaque cortical network	94	2390	50.9	111	0.51	0.88
<i>C. elegans</i> neuronal network	279	2990	21.4	137	0.79	0.60
airport flight connections	359	13 460	75.0	314	0.77	0.77

features (electronic supplementary material, table S1; see electronic supplementary material, figures S6 and S7 for other features).

In summary, only nonlinear growth can reproduce the analysed features of real-world networks (table 1). The NL_A model can account for CV and HRCC measures, while the DD and NL_P models can explain CV values and hub occurrence times.

3.5. Case study: are alterations of connectivity in preterm-birth related to nonlinear growth?

Finally, we have investigated the explanatory power of the NL model with regard to a recent study [24], where the brain structural connectome in very preterm-born adults was analysed. Counterintuitively, the very preterm brain exhibits a stronger rich-club architecture than the control brain, despite a decrease of white matter resources.

Interestingly, our model can account for this difference. We demonstrate this by simulating the growth of weighted networks exhibiting the rich-club property. Our simulations evolve according to two scenarios. In the control scenario, the network expands nonlinearly as proposed above. Moreover, the connections are weighted, i.e. connections are attributed a scalar value instead of a binary value. This weight (or connection strength) reflects the thickness of fibre tracts in brain networks (i.e. fibre tracts comprising many more fibres than others have higher weights). In our model, we assume that the weight of newly formed connections decreases linearly with developmental time, i.e. the early connections are the strongest. In the pathological scenario, the scalar is additionally decreased when 75% of the network size has been reached (i.e. referring to the time-point of premature birth). Hence, the number of connections is the same in both scenarios; however, the weights of connections developing after preterm birth are decreased in the second scenario (in accordance with the analysis results of Karolis *et al.* [24]).

For a valid comparison of the two scenarios, the normalized rich-club indices are computed by dividing the rich-club coefficient by the mean rich-club coefficient of 100 reference networks: $\phi(k)/\overline{\phi_r(k)}$, where $\phi(k)$ and $\phi_r(k)$ denote rich-club coefficients for grown and reference networks, respectively, for a given rich-club degree k . As shown in figure 5, the pathological networks indeed exhibit a stronger rich-club organization than control networks. The reason is that the simulated premature brain networks have comparably enhanced strength within rich clubs, as the weight factor is decreased for the connections that are formed late during development (i.e. connections of the last 25% nodes). Hence, our model is in agreement with the finding of Karolis *et al.* [24] that there are no topological connectivity differences between the two groups. More information on the growth model for the weighted networks is available in the electronic supplementary material.

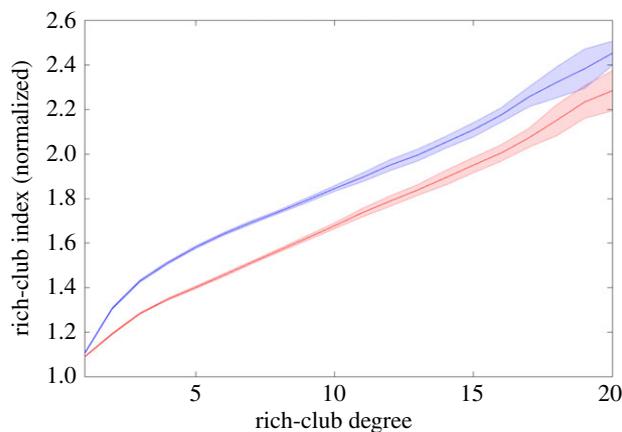


Figure 5. Normalized rich-club indices of control networks (red) and simulated, very preterm brain networks (blue). Shaded areas indicate standard deviation, as obtained from 20 curves per scenario. The x -axis indicates the rich club degree, i.e. the cut-off degree defining the rich-club. For the normalization of rich-club indices (y -axis), 100 reference networks preserving degree distribution were generated (hence, the number of nodes and node degree distribution do not differ in control, preterm and reference networks). These results demonstrate that the nonlinear growth model is in accordance with observations of increased rich-club organization in the structural connectome of very preterm-born adults (see also fig. 4 of [24]).

4. Discussion

We hypothesize that nonlinear growth explains hubs in many networks where nodes do not require detailed knowledge about nodes to which new nodes connect to (e.g. the degree of a node), and where information exchange is local. This is for instance the case for spatial networks where transmitting information over long distances can be prohibitively expensive. For example, during neural development the secretion and detection of diffusible substances are widely utilized processes for conveying critical information. However, the amount and detail of information that can be reliably transmitted in this way is limited. Our model solely requires that a newly added node can potentially form connections with the other nodes, and relies on less information than previous local models [7,25]. Along those lines, axons from one brain area can project to potential target areas based on guidance cues and/or genetically encoded growth mechanisms [26,27]. Hence, our model provides a framework where axonal connections do not need to rely on specific information about these targets, other than the well-established mechanisms required to connect with them. Crucially, the nonlinear growth model does not include space and is complementary to spatial models. As the model's requirements for the formation of connections are minimal, it is compatible with other models that include Euclidian distance [28] or properties of network connectivity [7].

Our work suggests that real-world networks substantially differ in terms of their hub-related network properties, depending on whether the principle of locality is a determinant of their development. Analyses of other real-world networks could support this hypothesis. Indeed, our model is unable to reproduce hub birth times in citation networks where links to spatially distant researchers nowadays do not involve extra costs (electronic supplementary material, figure S8).

In addition to biological and man-made networks, we particularly propose that nonlinear growth can account for the structural dynamics of neural networks during early (exponential) phases of brain development. In the mammalian brain, a single node could be a region of interest that comprises a large number (e.g. millions) of neurons. In more simple species, such as *C. elegans* or *Drosophila*, a node would correspond to only a few or single neurons. Indeed, studies on brain evolution and development support an exponential increase in the number of brain regions [9] and neurons [29–32]. Interestingly, in contrast with the model in Vértés *et al.* [33], nonlinear growth explains how rich-club organization arises without any access to non-local information. Moreover, our model is in agreement with observations of brain structure following neurodevelopmental disruption after very preterm birth; counterintuitively, rich-club organization is increased in such atypical scenarios [24]. Hence, the nonlinear growth model provides a general and mechanistic understanding of these findings.

The assessment of the growth models is conducted on a topological basis, hence functional differences between individual nodes are not accounted for. Also, we have focused on the development

of hub-related network features. In the future, additional growth mechanisms involving secondary mechanisms (removal of connections or node-specific connection formation) could be incorporated to establish small-worldness and modularity [28]. Such additional model complexity could be incorporated to generate networks matching the degree distribution itself, rather than the specific hub-related network measures which were assessed in this work. In addition, we only tested for accelerated growth where the number of new nodes at each step increases: it will be interesting to also test the effects of reductions in the growth rate, meaning scenarios where each step produces fewer nodes than for the previous steps. This aspect highlights that nonlinear growth, as linear growth, is limited by the maximum size of the network that can or should be sustained.

Our model for hub formation does not imply a scale-free degree distribution (electronic supplementary material, figure S9), which is an independent feature for some networks [17,34]. Hence, our model is well suited for scenarios where such scale-freeness is not established but hubs are nonetheless known to exist [35]. We emphasize that the nonlinear growth principle is complementary to many mechanistic explanations for how neural connectivity arises during development [17,36] and evolution [37].

Overall, nonlinear growth relies solely on locally available information (e.g. nodes do not require knowledge about the degree of connection partners), and so can provide a baseline benchmark for modelling network evolution that contains phases of exponential growth. We have demonstrated its explanatory power with respect to hub-related network features, in comparison with well-established classical models. Not all growth models that can generate hubs have been investigated in this study, because such an extensive study would entail an extremely high number of parameter optimizations. However, to the best of our knowledge, all alternative models assume that connecting nodes have access to more information than the nonlinear growth model relies upon (e.g. [38–40]). Therefore, the nonlinear growth model, either using relative or absolute numbers of edges from each new node, provides a general principle for hub development in certain spatially embedded and/or biological networks. Furthermore, we propose that the nonlinear growth model could be applied for the efficient design of spatial networks (e.g. power grids and supply networks), where robustness and information integration are crucial properties for network function [22].

Data accessibility. Our data are deposited at the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.6h8pm> [41].

Authors' contributions. R.B. conducted the computational work and produced the figures. M.K. aided in the design of the approach and in the data analysis. Both wrote the text.

Competing interests. The authors declare that they have no competing financial interests.

Funding. This work was supported by the Engineering and Physical Sciences Research Council of the United Kingdom (EP/K026992/1) as part of the Human Brain Development Project (<http://www.greenbrainproject.org/>). R.B. was also supported by the Medical Research Council of the UK (MR/N015037/1).

Acknowledgements. The authors wish to thank Sol Lim, Michael Pfeiffer, Peter Taylor, Yujiang Wang and Frédéric Zubler for stimulating discussions and helpful comments.

References

- Sporns O, Honey CJ, Kötter R. 2007 Identification and classification of hubs in brain networks. *PLoS ONE* **2**, e1049. (doi:10.1371/journal.pone.0001049)
- Variar S, Kaiser M. 2011 Neural development features: spatio-temporal development of the *Caenorhabditis elegans* neuronal network. *PLoS Comput. Biol.* **7**, e1001044. (doi:10.1371/journal.pcbi.1001044)
- Jeong H, Mason SP, Barabási AL, Oltvai ZN. 2001 Lethality and centrality in protein networks. *Nature* **411**, 41–42. (doi:10.1038/35075138)
- Goymer P. 2008 Network biology: why do we need hubs? *Nat. Rev. Genet.* **9**, 650–651. (doi:10.1038/nrg2450)
- Lloyd AL, May RM. 2001 How viruses spread among computers and people. *Science* **292**, 1316–1317. (doi:10.1126/science.1061076)
- Barabási AL, Albert R. 1999 Emergence of scaling in random networks. *Science* **286**, 509–512. (doi:10.1126/science.286.5439.509)
- Vázquez A, Flammini A, Maritan A, Vespignani A. 2002 Modeling of protein interaction networks. *Complexus* **1**, 38–44. (doi:10.1159/000067642)
- Monod J. 1949 The growth of bacterial cultures. *Annu. Rev. Microbiol.* **3**, 371–394. (doi:10.1146/annurev.mi.03.100149.002103)
- Ebbesson SO. 1980 The parcellation theory and its relation to interspecific variability in brain organization, evolutionary and ontogenetic development, and neuronal plasticity. *Cell Tissue Res.* **213**, 179–212. (doi:10.1007/BF00234781)
- Zhang GQ, Zhang GQ, Yang QF, Cheng SQ, Zhou T. 2008 Evolution of the Internet and its cores. *New J. Phys.* **10**, 123027. (doi:10.1088/1367-2630/10/12/123027)
- Rubinov M, Sporns O. 2010 Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* **52**, 1059–1069. (doi:10.1016/j.neuroimage.2009.10.003)
- Kim WK, Marcotte EM. 2008 Age-dependent evolution of the yeast protein interaction network suggests a limited role of gene duplication and divergence. *PLoS Comput. Biol.* **4**, e1000232. (doi:10.1371/journal.pcbi.1000232)
- Kaiser M, Hilgetag CC. 2006 Nonoptimal component placement, but short processing paths, due to long-distance projections in neural systems. *PLoS Comput. Biol.* **2**, e95. (doi:10.1371/journal.pcbi.0020095)
- Varshney LR, Chen BL, Paniagua E, Hall DH, Chklovskii DB. 2011 Structural properties of the *Caenorhabditis elegans* neuronal network. *PLoS Comput. Biol.* **7**, e1001066. (doi:10.1371/journal.pcbi.1001066)
- Marcelino J, Kaiser M. 2012 Critical paths in a metapopulation model of H1N1: efficiently delaying influenza spreading through flight cancellation. *PLoS Curr.* **4**, e4f8c9a2e1fca8. (doi:10.1371/4f8c9a2e1fca8)

16. Erdős P, Rényi A. 1959 On random graphs, I. *Publicationes Mathematicae (Debrecen)* **6**, 290–297.
17. Kaiser M, Hilgetag CC. 2004 Spatial growth of real-world networks. *Phys. Rev. E* **69**, 036103. (doi:10.1103/PhysRevE.69.036103)
18. Colizza V, Flammini A, Serrano MA, Vespignani A. 2006 Detecting rich-club ordering in complex networks. *Nat. Phys.* **2**, 110–115. (doi:10.1038/nphys209)
19. Towilson EK, Vértes PE, Ahnert SE, Schafer WR, Bullmore ET. 2013 The rich club of the *C. elegans* neuronal connectome. *J. Neurosci.* **33**, 6380–6387. (doi:10.1523/JNEUROSCI.3784-12.2013)
20. de Reus MA, van den Heuvel MP. 2013 Rich club organization and intermodule communication in the cat connectome. *J. Neurosci.* **33**, 12 929–12 939. (doi:10.1523/JNEUROSCI.1448-13.2013)
21. Harriger L, Van den Heuvel MP, Sporns O. 2012 Rich club organization of macaque cerebral cortex and its role in network communication. *PLoS ONE* **7**, e46497. (doi:10.1371/journal.pone.0046497)
22. van den Heuvel MP, Sporns O. 2011 Rich-club organization of the human connectome. *J. Neurosci.* **31**, 15 775–15 786. (doi:10.1523/JNEUROSCI.3539-11.2011)
23. van den Heuvel MP, Sporns O. 2013 Network hubs in the human brain. *Trends Cogn. Sci.* **17**, 683–696. (doi:10.1016/j.tics.2013.09.012)
24. Karolis VR *et al.* 2016 Reinforcement of the brain's rich-club architecture following early neurodevelopmental disruption caused by very preterm birth. *Cereb. Cortex* **26**, 1322–1335. (doi:10.1093/cercor/bhv305)
25. Wang LN, Guo JL, Yang HX, Zhou T. 2009 Local preferential attachment model for hierarchical networks. *Phys. A Stat. Mech. Appl* **388**, 1713–1720. (doi:10.1016/j.physa.2008.12.028)
26. Dodd J, Jessell TM. 1988 Axon guidance and the patterning of neuronal projections in vertebrates. *Science* **242**, 692–699. (doi:10.1126/science.3055291)
27. Lumsden AG, Dodd J. 1988 Chemotropic guidance of developing axons in the mammalian central nervous system. *Nature* **336**, 22. (doi:10.1038/336022a0)
28. Nisbach F, Kaiser M. 2007 Developmental time windows for spatial growth generate multiple-cluster small-world networks. *Eur. Phys. J. B* **58**, 185–191. (doi:10.1140/epjb/e2007-00214-4)
29. Rakic P. 1995 A small step for the cell, a giant leap for mankind: a hypothesis of neocortical expansion during evolution. *Trends Neurosci.* **18**, 383–388. (doi:10.1016/0166-2236(95)93934-P)
30. Caviness V, Takahashi T, Nowakowski R. 1995 Numbers, time and neocortical neuronogenesis: a general developmental and evolutionary model. *Trends Neurosci.* **18**, 379–383. (doi:10.1016/0166-2236(95)93933-0)
31. White T, Su S, Schmidt M, Kao CY, Sapiro G. 2010 The development of gyrification in childhood and adolescence. *Brain Cogn.* **72**, 36–45. (doi:10.1016/j.bandc.2009.10.009)
32. Butts T, Green MJ, Wingate RJ. 2014 Development of the cerebellum: simple steps to make a 'little brain'. *Development* **141**, 4031–4041. (doi:10.1242/dev.106559)
33. Vértes PE, Alexander-Bloch A, Bullmore ET. 2014 Generative models of rich clubs in Hebbian neuronal networks and large-scale human brain networks. *Phil. Trans. R. Soc. B* **369**, 20130531. (doi:10.1098/rstb.2013.0531)
34. Markov N *et al.* 2014 A weighted and directed interareal connectivity matrix for macaque cerebral cortex. *Cereb. Cortex* **24**, 17–36. (doi:10.1093/cercor/bhs270)
35. Stumpf MP, Porter MA. 2012 Critical truths about power laws. *Science* **335**, 665–666. (doi:10.1126/science.1216142)
36. Bauer R, Zubler F, Pfister S, Hauri A, Pfeiffer M, Muir DR, Douglas RJ. 2014 Developmental self-construction and -configuration of functional neocortical neuronal networks. *PLoS Comput. Biol.* **10**, e1003994. (doi:10.1371/journal.pcbi.1003994)
37. Chakraborty M, Jarvis ED. 2015 Brain evolution by brain pathway duplication. *Phil. Trans. R. Soc. B* **370**, 20150056. (doi:10.1098/rstb.2015.0056)
38. Dorogovtsev SN, Mendes JFF, Samukhin AN. 2000 Structure of growing networks with preferential linking. *Phys. Rev. Lett.* **85**, 4633–4636. (doi:10.1103/PhysRevLett.85.4633)
39. Saramäki J, Kaski K. 2004 Scale-free networks generated by random walkers. *Phys. A Stat. Mech. Appl.* **341**, 80–86. (doi:10.1016/j.physa.2004.04.110)
40. Gabel A, Kraivsky P, Redner S. 2014 Highly dispersed networks generated by enhanced redirection. *J. Stat. Mech. Theory Exp.* **2014**, P04009. (doi:10.1088/1742-5468/2014/04/P04009)
41. Bauer R, Kaiser M. 2017 Data from: Nonlinear growth: an origin of hub organization in complex networks. Dryad Digital Repository. (<http://dx.doi.org/10.5061/dryad.6h8pm>)